

WHAT IS CLAIMED IS:

1. A method of operating a server cluster including a set of server devices each connected to a local area network, comprising:

5

preventing access to a selected server's memory by other servers on the server cluster when the selected server is powered up;

deactivating the selected server responsive to a decrease in server cluster traffic;

10

responsive to deactivating the selected server, permitting the other servers on the cluster to access at least a portion of the selected server's memory; and

15

responsive to a request received by one of the other servers for a file stored in the selected server's file cache, retrieving the file from the selected server's file cache over the local area network.

20

2. The method of claim 1, wherein permitting the other servers to access at least a portion of the selected server's memory includes broadcasting a directory of the selected server's file cache contents to the other servers before powering down the selected server.

3. The method of claim 2, further comprising, prior to powering down the selected server, processing any pending client requests on the selected server.

25

4. The method of claim 1, further comprising, responsive to an increase in network traffic:

broadcasting a message to each of the other server devices on the server cluster that the selected server is to be activated;

responsive to the activation message, preventing the other servers from accessing the system memory of the selected server; and

activating the selected server.

5

5. The method of claim 4, wherein preventing the other servers from accessing the system memory of the selected server includes deleting a directory of the selected server's file cache contents from the memories of the other servers.

10

6. The method of claim 1, wherein deactivating the selected server includes transitioning the selected server's processors to a low power state while maintaining power to the selected server's NIC and system memory.

15

7. The method of claim 1, wherein retrieving the file from the selected server's file cache includes initiating a direct memory access of the selected server's system memory from the selected server's NIC.

8. A server device suitable for use in a data processing network comprising:

20

at least one processor;

a system memory accessible to the processor via a system bus;

25

a network interface card (NIC), including a NIC controller and memory, connected to the system bus and providing a connection to the local area network;

server code means for deactivating the processor;

NIC code means for preventing access to the server device's memory by other servers on the server cluster when the server device is activated; and

NIC code means for enabling the other servers to retrieve a file from the system memory of the server device when the server device is deactivated.

9. The server of claim 8, wherein the code means for enabling the other servers to retrieve a file from the server device includes code means for broadcasting a directory of the server device's file cache contents prior to deactivation.

10. The server of claim 8, wherein the code means for deactivating the server includes code means for powering down the server's processors while maintaining power to the server's NIC and system memory.

11. The server of claim 8, wherein the NIC is configured to access the server system memory directly without invoking the server processor.

12. The server of claim 8, wherein the NIC is a PCI compliant NIC.

13. A data processing network including a set of interconnected server devices, each server device comprising:

at least one processor;

a system memory accessible to the processor via a system bus;

a network interface card (NIC), including a NIC controller and memory, connected to the system bus and providing a connection to the local area network;

server code means for deactivating the processor;

NIC code means for preventing access to the server device's memory by other servers on the server cluster when the server device is activated; and

5

NIC code means for enabling the other servers to retrieve a file from the system memory of the server device when the server device is deactivated.

10 14. The network of claim 13, wherein the code means for enabling the other servers to retrieve a file from the server device includes code means for broadcasting a directory of the server device's file cache contents prior to deactivation.

15 15. The network of claim 13, wherein the code means for deactivating the server includes code means for powering down the server's processors while maintaining power to the server's NIC, peripheral bus, and system memory.

16. The network of claim 13, wherein the NIC is configured to access the server system memory directly without invoking the server processor.

20 17. The network of claim 13, wherein the NIC is a PCI compliant NIC.

18. The network of claim 13, further comprising network code means for dynamically adjusting the number of deactivated servers in the network responsive to variations in network traffic.

25 19. The network of claim 18, further comprising network code means for directing client to servers of the network based at least in part on the requested content.

20. The network of claim 13, wherein each of the server devices includes code means for retrieving a requested file from the file cache of a deactivated server on the server cluster responsive to determining that the requested file is in the deactivated server's file cache.

FOR PUBLICATION